6. Use of Expert Judgment

6.1 INTRODUCTION

6.1.1 Background

In 40 CFR part 194, EPA states that expert judgment should only be permitted in situations where data are not reasonably obtainable by collection or experimentation. EPA requires that compliance applications clearly identify all instances in which judgment is used and the experts involved. Documentation must be included which describes the process for expert judgment elicitation, the results of expert elicitation, and the reasoning behind those results. Documentation of interviews used to elicit judgments from experts, deliberations and formal interactions among experts, background information provided to experts, and the questions or issues presented for elicitation of expert judgment are also requested.

Although the Agency has not specified any particular methods for expert judgment elicitation, the Agency has included some restrictions and guidelines for the selection of individuals as experts in the 40 CFR part 194 criteria. These include prohibitions on: selecting individuals who are members of the team of investigators requesting the judgment or the team of investigators who will use the judgment; selecting individuals who maintain a supervisory role or who are supervised by those who will utilize the judgment; and selecting a membership of which less than two-thirds consists of individuals who are not employed directly or indirectly by DOE (unless it can be shown that this is impracticable because of a lack or unavailability of qualified independent experts, in which case at least one-third of the membership must be non-DOE personnel). Compliance certification applications must provide information which demonstrates that the expertise of any individuals and the panel, as a whole, involved in expert judgment is consistent with the level of knowledge required by the questions or issue presented to that individual and the panel.

Additionally, EPA requires that at least five individuals be used in any expert elicitation process, unless a lack or unavailability of experts can be demonstrated and documented. Also, any compliance certification application should include a discussion explaining the relationship between the information presented, the questions asked, the judgment of any expert panel or individual, and the purpose for which the expert judgment is being used.

EPA requires that a minimum of five persons form an expert panel so that the elicited results are representative of diverse viewpoint. This should result in a more informed and objective process. However, an expert elicitation could be conducted with fewer than five individuals in the event that there is a lack or unavailability of potential experts, provided that a rational is stated. §194.26 of the final rule states this restriction:

> At least five individuals shall be used in any expert elicitation process, unless there is a lack or unavailability of experts and a documented rational is provided that explains why fewer than five individuals were selected.

It is essential that any expert panel member should be free from conflict of interest. Accordingly, two-thirds of the members of any panel should not be employed by DOE, directly or indirectly. This restriction does not extend to those persons who receive funding from the Department in those instances in which such funding is for activities not related to WIPP (such as university professors). Expert panels may include persons employed by the State of New Mexico Environmental Evaluation Group assuming that their expertise can be demonstrated to be adequate for the elicitation. Compliance applications must demonstrate this expertise, and EPA's judgment on the adequacy of this demonstration will be used in making the decision on the issuance of certification.

Finally, EPA proposes that the elicitation process provide the public an opportunity for presentation of scientific and technical views to the experts.

The Background Information Document (BID) for the 1993 amendments to 40 CFR part 191 notes that "It is generally accepted that the use of expert judgment is required in the process of evaluating the long-term containment potential of a geologic waste disposal facility" (EPA93). In this context, the term "expert judgment" refers to a very structured, formalized process involving panels of experts. However, expert judgment may also be applied by an individual charged with making a determination on a given situation.

DOE is using both expert panel and individual investigator judgment to support the WIPP performance assessment. In some instances, an expert panel may be convened and opinion elicited using a highly structured, formal approach. In other cases, a single principal investigator may be asked to supply an estimate of a parameter where a limited amount of experimental data is available and also provide an estimated probability distribution function for that parameter. The principal investigator may also be asked to define the probability distribution function for a parameter where considerable experimental data are available, but

which still must be interpreted.

6.1.2 NRC Publications on the Expert Judgment Process

NUREG/CR-5424 (NRC91) notes that the process by which expert judgment is elicited will vary depending on the particular situation. NUREG/CR-5424 lists the following factors that may affect how the judgment can best be gathered:

- The type of information needed from the experts (answers only or ancillary expert data)
- The form in which the expert's answers are needed for input into a model
- The number of experts available
- The interaction desired among the experts
- The difficulty of setting up the problems
- The amount of time and study needed by the experts to provide judgments
- The time and resources available to the study
- The methodological preferences of the interviewer or knowledge engineer, analyst, funder, and experts

Among the ways in which elicitation processes may differ include the degree to which the experts interact, the structure imposed on the process, the number of meetings, whether the expert's reasoning is requested or not, whether the expert judgment undergoes some translation in a model and is returned to the experts for the next step, and whether all or some of the elicitation is conducted in person, by mail, or by telephone.

NUREG/CR-5424 states that despite these variations, there are only three basic elicitation situations and a general sequence of steps. The three basic situations are as follows:

> Individual interviews - where one expert is interviewed in a private, usually face-to-face situation, by an interviewer or knowledge engineer (a person who, in addition to interviewing, represents and enters the expert knowledge into a computer system). This situation permits obtaining in-depth data from experts, such as on their means of solving the problem, without distracting or influencing them with other experts.

Interactive groups - where the experts are in a face-to-face situation with both one another and a session moderator when they give their opinion. The degree of structure may vary from totally unstructured to carefully choreographed as to when the experts present their views and when there is open discussion.

Delphi - where the experts give their judgments to a moderator, in isolation from one another. The moderator makes the judgments anonymous, redistributes them to the experts, and allows them to revise their previous judgments. If desired, the iterations can be continued to the point where consensus is achieved. This process is intended to counter some of the biasing effects of interaction.

The general sequence of steps in the elicitation process are described below (NRC91):

1. Selection of the question areas and particular questions
2. Refining of the questions
3. Selection and motivation of the experts
4. Selection of the components (building blocks) of elicitation
5. Designing and tailoring of the components of elicitation to fit the application
6. Practicing the elicitation and training the in-house personnel
7. Eliciting and documenting expert judgments (answers, and/or ancillary information)

NUREG/CR-5411,"Elicitation and Use of Expert Judgement in Performance Assessment for High-Level Radioactive Waste Repositories" (NRC90b), indicates five areas of performance assessment of high-level waste repositories for which the benefits of a formal expert judgment process may be warranted:

-scenario development and screening
-model development
-parameter estimation
-data collection and experimentation (information gathering)
-strategic repository decisions

40 CFR part 194 does not require that expert judgment be applied to any one area in particular, but leaves this choice up to DOE, subject to the restrictions in the final rule.

A well-documented application of the formal use of expert judgment is the U.S. Nuclear Regulatory Commission's study, NUREG/CR-1150, "Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants" (NRC90). That study was undertaken to provide a risk perspective for the radioactive release resulting from a core meltdown (see section 6.2.1.1 for

further discussion of NUREG/CR-1150).

6.2 EXAMPLES OF THE USE OF EXPERT JUDGMENT AT FACILITIES OTHER THAN THE WIPP

Expert judgment has been used in various scientific forums not related to the WIPP. Four reports on use of expert judgment are reviewed in this section: the NRC's application in severe accident risk assessment; the Electric Power Research Institute's use in a probabilistic seismic hazard analysis; the United Kingdom's use in risk assessment of radioactive waste disposal; and the European Space Agency's examination of expert judgment for risk assessment in space programs.

6.2.1 Nuclear Regulatory Commission

6.2.1.1 NUREG/CR-1150

A well-documented use of expert judgment in the area of nuclear reactor safety is presented in NUREG/CR-1150. The report summarizes an assessment of the risks from severe accidents in five commercial nuclear power plants in the U.S. The risks were measured in a number of ways, including: the estimated frequencies of core damage accidents from internally initiated accidents and externally initiated accidents for two of the plants; the performance of containment structures under severe accident loadings; the potential magnitude of radionuclide releases and offsite consequences of such accidents; and the overall risk (the product of accident frequencies and consequences) (NRC90).

The report notes that the risk analysis of severe reactor accidents inherently involves the consideration of parameters for which little or no experiential data exist. Expert judgment was used to supplement and interpret the available data. The principal steps used in the formal elicitation of expert judgment for the NUREG/CR-1150 study are shown schematically in Figure 6-1 and discussed briefly below:

- Selection of issues - The parameters considered were restricted to those with the largest uncertainties, expected to be the most important to risk, and for which widely accepted data were not available.

- Selection of experts - Seven panels of experts were assembled to consider the various sets of principal issues. The experts were selected on the basis of their recognized expertise in the issue areas. Representatives from the nuclear industry, the NRC and its contractors, and academia were assigned to each

panel to ensure a balance of perspectives.

- Training in elicitation methods - Both the experts and analysis team members received training from specialists in decision analysis. The team members were trained in elicitation methods so that they would be proficient and consistent in their elicitation. The experts' training included an introduction to the elicitation and analysis methods, to the psychological aspects of probability estimation (e.g., the tendency to be overly confident in the estimation of probabilities), and to probability estimation.

- Presentation and review of issues - Presentations were made to each panel on the set of issues to be considered, the definition of each issue, and relevant data on the issues. Also, for the initial meeting, researchers, plant representatives, and interested parties were invited to present their perspectives on the issues to the experts. NUREG/CR-1150 notes that frequently these presentations took several days.

- Preparation of expert analyses - Following the initial meeting in which the issues were presented, the experts were given time (from 1 to 4 months) to prepare their analyses. During this period, several panels met to exchange information and ideas. In some cases, panels were briefed by the project staff on the results from other panels to provide the most current data.

- Expert review and discussion - After the experts had completed their analyses a final meeting was held in which each expert discussed the methods he or she used to analyze the issue. NUREG/CR-1150 states that while these discussions frequently led to modifications of the preliminary judgments of individual experts, the experts' actual judgments were not discussed in the meeting because group dynamics can cause people to unconsciously alter their judgments in the desire to conform.

- Elicitation of experts - Following the panel discussions, each expert's judgments were solicited. The elicitations were done privately with one expert at a time so that the discussions could be performed in depth and so that an expert's judgments would not be adversely influenced by the others.

- Composition and aggregation of judgments - The analysis staff composed probability distributions for each expert's judgments, and then aggregated the individual judgments to provide a single composite judgment for each issue. NUREG/CR-1150 notes that each expert's opinion was weighted equally in the aggregation, based on findings in previous studies that this method performs best.

- Review by experts - Each expert's probability distribution and associated documentation developed by the analysis staff were reviewed by that expert. The purpose of this review was to ensure that potential misunderstandings were identified and corrected and that the issue documentation properly reflected the

judgments of the expert.

Figure 6-1. Principal Steps in NUREG/CR-1150 Expert Elicitation Process

Results of the elicitation are presented in NUREG/CR-1150, and documented in detail in two separate reports (NRC89, NRC90a). While members of each panel are identified in NUREG/CR-1150, specific judgments are presented anonymously (e.g., the members are identified only as Expert A or Expert B).

6.2.1.2 Yucca Mountain Climate Study

More recently, the use of expert judgment elicitation was examined in a study to predict future climate in the vicinity of Yucca Mountain, Nevada, the site currently undergoing characterization for DOE's high-level waste repository. A report documenting that study was presented at the Fifth Annual International Conference on High-Level Radioactive Waste Management in Las Vegas, Nevada, in May 1994 (DeW94).

The expert elicitation procedure used consisted of the following 11 steps, adapted from NUREG/CR-5411 (NRC90b):

1. Determine the objectives and goals of the study
2. Recruit the specialists (experts)
3. Identify the issues and information needs
4. Provide initial data to the specialists
5. Conduct the elicitation training session
6. Discuss and refine the issues
7. Provide a multi-week study period
8. Conduct the elicitation
9. Provide post-elicitation feedback
10. Aggregate the experts' judgments (if required)
11. Document the process

The elicitation team (those persons who would be responsible for conducting the elicitation) and the expert panel were recruited concurrently with the development of the issue statement. Nominations for expert panelists were formally requested from climatology/geography associated societies and organizations. A formal peer-ranking based selection process resulted in five final panel members, drawn from 42 nominees.

An initial meeting was held, with three goals: (1) to orient the experts; (2) to refine the initial issue statement; and (3) to conduct elicitation training. The experts received background information on the proposed repository system, the current and past climate in the Yucca

Mountain vicinity, and the NRC's performance assessment program. Extensive training was provided on probability elicitation, including the interpretation of subjective probabilities, methods for generating subjective probabilities, and possible biases in the judgment process. The experts refined the initial issue statement and generated a list of factors and assumptions that would be considered by the group.

The experts had one month between the initial meeting and the individual elicitation to review any relevant literature, run models, or otherwise prepare for the elicitation, including providing a position paper. The experts had access to each other for consultation to exchange data or clarify information, but they prepared their positions independently. The actual elicitations were conducted individually to obtain the independent judgment of each expert.

DeW94 concluded that while each of the steps used in the elicitation process influences the outcome, four points are critical:

(1) the process of recruiting the experts should be formal and as unbiased as possible;

(2) the credentials of the experts enhance the credibility of the elicitation, and their ability to communicate their reasoning is a primary determinant of the quality of the results;

(3) the conduct of the elicitation sessions themselves is extremely important and should be well-planned and practiced ahead of time;

(4) concise and thorough documentation of the process including recording of the elicitation sessions, as well as the results, differentiates between most informal and formal expert judgment efforts and is essential in any formal expert elicitation project.

6.2.2 Electric Power Research Institute

Expert judgment was used in a 1983 effort by the Electric Power Research Institute (EPRI) and the 42 utilities in the Seismicity Owners Group (SOG) to develop a methodology for assessing the seismic hazard at nuclear power plant sites. The results are documented in the report *Seismic Hazard Methodology for the Central and Eastern United States* (EPR88).

In the EPRI study, earth science expertise was provided by teams formed specifically to promote interaction among different disciplines (geology, geophysics, seismology) and thus avoid an overly narrow disciplinary focus. Work was conducted through a series of workshops. Participating Earth Science Teams were required to identify and document specific factors that could be used to evaluate the activity of tectonic features. For each feature, each team was also required to assess the extent to which those factors were exhibited. This established a distinction between scientific uncertainty (uncertainty in the relationship between tectonic activity and physical characteristics) and information uncertainty (the extent to which any particular feature exhibits any given characteristics).

Six Earth Science Teams were formed to prepare and interpret input to the seismic hazard analysis. EPRI used the team approach to achieve the interdisciplinary expertise needed to evaluate various data sets and tectonic processes on a national scale. Team personnel were chosen to strike a balance between, first, academic and applied experience, and second, regional expertise. The stated overall aim was to minimize interpretation bias.

This study was accomplished through a series of seven workshops. The first workshop defined data needs for the program. Workshops 2 through 7 were structured in pairs to accomplish interpretations of tectonic stress regime, tectonic framework and seismic sources and source seismicity parameters. Procedures were explored in depth during each workshop to establish a common understanding among participants of the state of knowledge about processes and the relative value of available data for making interpretations.

The Earth Science Teams proceeded with this information and their personal expertise to develop their individual interpretations. Interpretations were shared among program participants at the second workshop of each pair. Each team shared the rationale and the strength of theory and data supporting its interpretations. EPRI notes that although this team-to-team interaction was desired, no effort was made to force a consensus interpretation among teams on any element. Teams were asked to reach internal consensus on all interpretations within a team.

EPRI stated that, with this approach, "..it is believed that uncertainty resulting from incomplete understanding of tectonic processes has been captured. The estimated uncertainty in hazard results ... reflects the state of the scientific community's uncertainty about earthquakes causes and processes in the central and eastern United States." (EPR88)

6.2.3 Other Countries

6.2.3.1 United Kingdom

A 1992 report commissioned in the United Kingdom by Her Majesty's Inspectorate of Pollution, Department of the Environment (DoE) examined procedures for the elicitation of expert judgments in probabilistic risk analysis of radioactive waste repositories (WAT92). The report concluded that "Expert judgment is necessary for the measurement of uncertainty about input parameters, since for many such parameters no frequency data are available."

WAT92 described several sources of bias that can influence judgments. The first noted was that of *availability*. This is based on the observation that people are often influenced by the ease with which they can remember the occurrence of similar events. An example was cited in which a group of well educated people in the United States were told that about 50,000 people a year die in traffic accidents in the United States. They were then asked how many people die from a long list of other causes, including common ailments such as heart disease and rare ones such as smallpox vaccination. It was found that rare causes were overestimated while common causes were underestimated. The explanation was that deaths from a rare cause such as botulism are widely reported, but people commonly hear about deaths from common causes such as stroke only when someone known to them dies in this way. Because all cases of rare causes are available, but not all cases of common causes, rare causes are overestimated.

A second possible source of bias is *representativeness*. This is based on the premise that, for example, when people attempt to assess the probability that an individual belongs to a particular class, on the basis of limited information, they judge the extent to which that information suggests the individual is typical of the class, ignoring the underlying frequency. In terms of a repository, an expert asked for the probability that the porosity of a rock formation was greater than a certain figure might base his judgment on the extent to which observable characteristics of the rock samples suggested that it was of a particular type of known porosity, independent of the known distribution of different types of rock in the area. This emphasizes the need for a published account of the reasons supporting a probability judgment.

The third potential source of bias presented was due to *anchoring and adjustment*. This recognizes that a natural starting point for making a judgment may be chosen and the judgment modified away from the initial position, but typically not far enough. Again using the example of estimating the porosity of a rock, an expert might base judgment on a different rock, for which the porosity was well known, but fail to adjust the estimate away from this properly.

WAT92 also examined protocols for eliciting probabilities. It noted that the protocol developed at the Stanford Research Institute, and referred to as the SRI protocol, has been widely accepted. The SRI protocol has five stages.

> In the first stage, the analyst *motivates* the person whose probabilities are to be elicited. The first step in this process is to ensure that the expert understands the nature and purpose of the analysis, and how the probabilities elicited will be used in the analysis. In the second step, the analyst helps the expert to explore for possible motivational biases, e.g., if the expert desires a low value of a variable, either because he thinks that this would be consistent with what his superior is expecting him to say, or because he wishes it were true, then such views should be discovered, if possible, and the expert encouraged to account for them in the elicitation task.
>
> The next stage in the protocol is the *structuring* phase. In this phase, the goal is to make absolutely clear the definition of the variable for which the probability distribution is being elicited. This should also include the exploration of assumptions about the state of the world (e.g., in eliciting probability distributions for the porosity of rock, assumptions made about physical variables such as temperature and pressure which could affect porosity should be clearly defined).
>
> The *conditioning* phase in the protocol is used to establish the data and arguments which the expert is going to use to make judgments, and to cope with any identified biases. Once the available data sets are listed, the expert is encouraged to consider other possible ways of thinking about the variable, for example, focusing on scenarios that might lead to extreme outcomes.
>
> Numerical representation begins in the fourth, or *encoding*, phase. The techniques for encoding can be categorized according to whether the probabilities are inferred indirectly from the expert's judgments, or directly by asking the expert to respond with a probability for a given event, or, in the case of continuous variables, for a value such that the cumulative probability is equal to a given value. WAT92 notes that there is no consensus that any one method is better than another, but that there is a tendency for experts to be overconfident, i.e., they fail to spread their uncertainty sufficiently.

The final stage involves *verifying* that the numerical representations of uncertainty properly support the expert's opinions. Assuming that no recognized protocols were identified, the analyst must determine how best to conduct this process.

WAT92 does not reach a conclusion that any one protocol is superior, but it notes several important considerations. First, the most important requirement for success is to devote adequate time and effort to the process. Second, the experts should receive some general introduction on the nature of elicitation, and what is known from psychology about measuring perceptions of uncertainty. Third, the subject matter about which the judgments are sought should be at the heart of the process. The analyst should have some knowledge of the subject matter, and the experts should be encouraged to produce carefully reasoned arguments to support his judgments. Fourth, throughout the process care must be taken to ensure clarity and investigate unstated assumptions. Fifth, the encoding process should follow a generally accepted technique.

6.2.3.2 European Space Agency

In February 1990 the European Space Agency (ESA) released a report entitled *The Use of Expert Judgment in Risk Assessment* (COO90). The report provides the results of an examination of expert judgment application. The examination included (1) a survey identifying and studying the different and most important methods for the use of expert judgment; (2) a survey of application of expert judgment data in industries, research institutes and other organizations; and (3) development and evaluation of methodologies for expert judgment application.

The major conclusion of the ESA research was that "The introduction of valid and effective procedures for the use of expert judgment in risk assessment is a non-trivial, but worthwhile task."

Nine phases in the procedures for using expert judgment were identified:

1. Problem identification phase
2. Expert identification phase
3. Expert choice phase
4. Question formulation phase
5. Seed variable choice

6. Elicitation phase
7. Combination phase
8. Discrepancy analysis and feedback to experts
9. Documentation and communication phase

While the order and format may be somewhat different, these phases are approximately equivalent in content to the steps previously described in section 6.1.3, with the exception of step five, seed variable choice. The ESA study attempted to assess the quality of the experts' judgments before their use, and establish a basis for calibrating the judgments. This was done by eliciting the experts' judgment for quantities, known as seed or calibration variables, that were known to the analyst but not to the experts.

One conclusion of the ESA study was that elicitation should be conducted individually, i.e., not in a group setting. The report notes "The advantage of group meetings is that experts can discuss together with the analyst the interpretation of all the questions so encouraging a shared understanding of their meaning. The disadvantage is that group processes naturally suppress the spread of opinion and lead to underestimating uncertainty. The balance lies we feel with the group not meeting." The report states that, if it is decided that the group should meet, they meet to discuss the questions to be answered and separate before any elicitation of the likelihoods takes place. This is consistent with the techniques used in the NUREG/CR-1150 study, where groups met for presentation and discussion of the issues, but the individual elicitations were conducted in private.

The ESA report also identified the need for the generation of an audit trail in the use of expert judgment data to permit other analysts to repeat and check the information. The report considered the issue of whether the experts should be protected with anonymity, and if so, to what extent. Opinions were divided, but the report concluded the audit trail should be such that "in circumstances of sufficient gravity" the ESA (as the client soliciting the expert judgments) could identify all experts and their judgments.

6.3 REFERENCES

COO90 R. Cooke, S. French, and J. van Steen, "The Use of Expert Judgment in Risk Assessment," European Space Agency Reference CR(P) 3183, February 1990.

DeW94 A. DeWispelare, M. P. Miklas, A. B. Gureghian, R. T. Clemson, L. T. Herren, and J. R. Park, "The Use of Expert Judgement Elicitation to Predict Future Climate for the Yucca Mountain Nevada Vicinity," Proceedings of the Fifth Annual International Conference on High-Level Radioactive Waste Management, Las Vegas, Nevada, pp. 1614-1621, May 22-26, 1994.

EPA93 U. S. Environmental Protection Agency, "Background Information Document for Amendments to 40 CFR part 191, Environmental Standards for the Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes," EPA 402-R-93-073, November 1993.

EPR88 Electric Power Research Institute, "Seismic Hazard Methodology for the Central and Eastern United States," Volume 1, Part 2: Methodology (Revision 1), Final Report, November 1988.

NRC89 T.A. Wheeler et al., "Analysis of Core Damage Frequency from Internal Events: Expert Judgment Elicitation," Sandia National Laboratories, NUREG/CR-4550, Vol. 2, SAND86-2084, April 1989.

NRC90 NUREG/CR-1150, Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants," Final Summary Report, December 1990.

NRC90a F.T. Harper et al., "Evaluation of Severe Accident Risks: Quantification of Major Input Parameters," Sandia National Laboratories, NUREG/CR-4551, Vol.2, Revision 1, SAND86-1309, December 1990.

NRC90b NUREG/CR-5411, Elicitation and Use of Expert Judgement in Performance Assessment for High-Level Radioactive Waste Repositories, May 1990.

NRC91 NUREG/CR-5424, Eliciting and Analyzing Expert Judgment, A Practical Guide, March 1991.

WAT92 Watson, S.R., "Procedures of the Elicitation of Expert Judgements in the Probabilistic Risk Analysis of Radioactive Waste Repositories: an Overview," Research commissioned by Her Majesty's Inspectorate of Pollution, Department of the Environment, December 1992.